

NIST Research Data Framework (RDaF)

Dr. Robert Hanisch

Director, Office of Data and Informatics (ODI)

Material Measurement Laboratory

US National Institute of Standards and Technology

About NIST and ODI

- The National Institute of Standards and Technology is a federal agency under the US Department of Commerce
 - Known as the National Bureau of Standards until 1988, originally founded in 1901
- Non-regulatory
- State of the art in measurement science and technology
- US National Metrology Institute, amongst network of 103 NMIs globally organized under the Bureau International des Poids et Mesures (BIPM, or International Bureau of Weights and Measures), Paris
- ~5,000 staff at NIST (Gaithersburg, Maryland headquarters; Boulder, Colorado; Charleston, South Carolina; Brookhaven National Laboratory)
- 6 major research laboratories
 - Material Measurement Laboratory
 - Office of Data and Informatics (15 people)



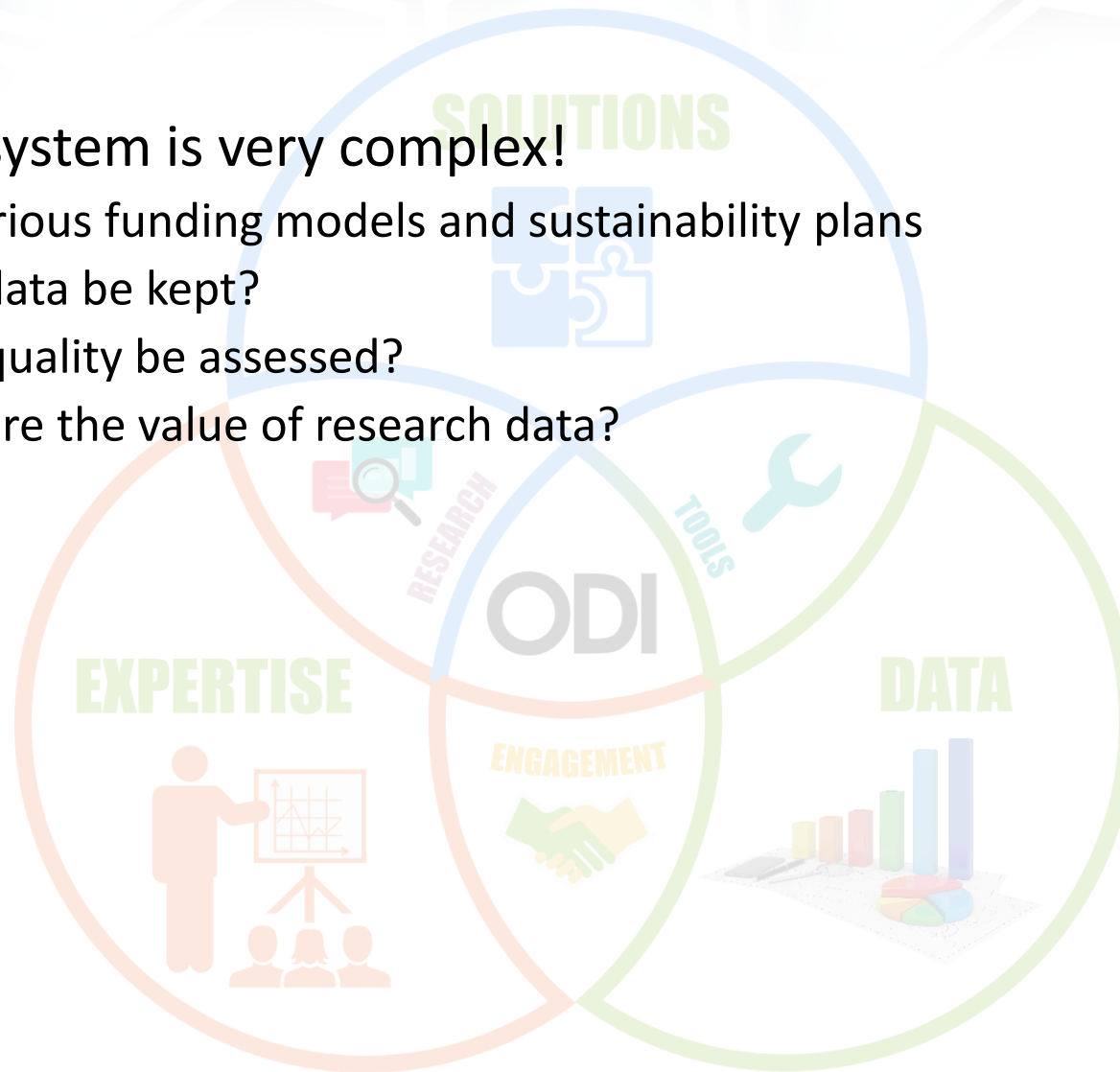
What is a Research Data Framework?

- A map of the research data space: who, what, where, why, when?
- A dynamic guide for the various stakeholders in research data to understand best practices for research data management and dissemination
- A resource for understanding costs, benefits, and risks associated with research data management
- A consensus document based on inputs and conversations amongst the stakeholders in research data



Why a Research Data Framework?

- Research data ecosystem is very complex!
 - Lots of players, various funding models and sustainability plans
 - How long should data be kept?
 - How should data quality be assessed?
 - How do we measure the value of research data?

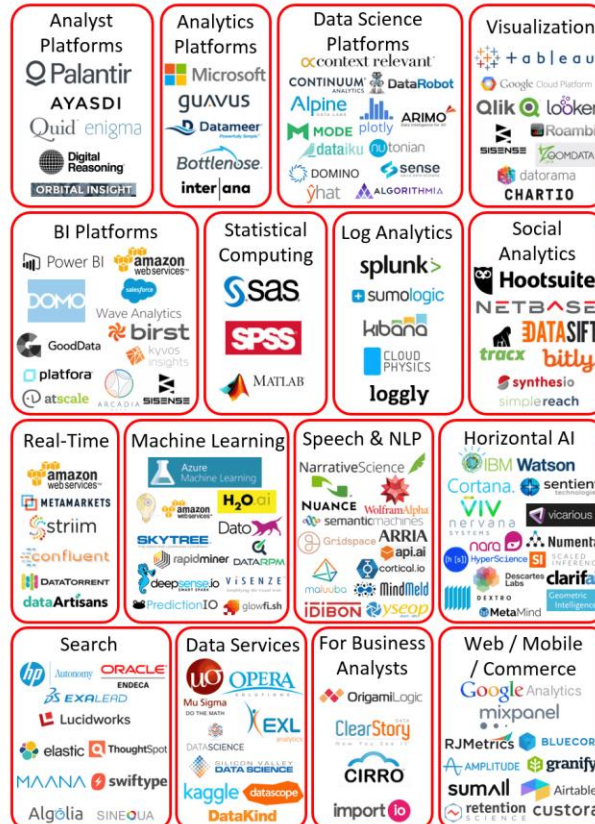


Big Data Landscape 2016 (Version 3.0)

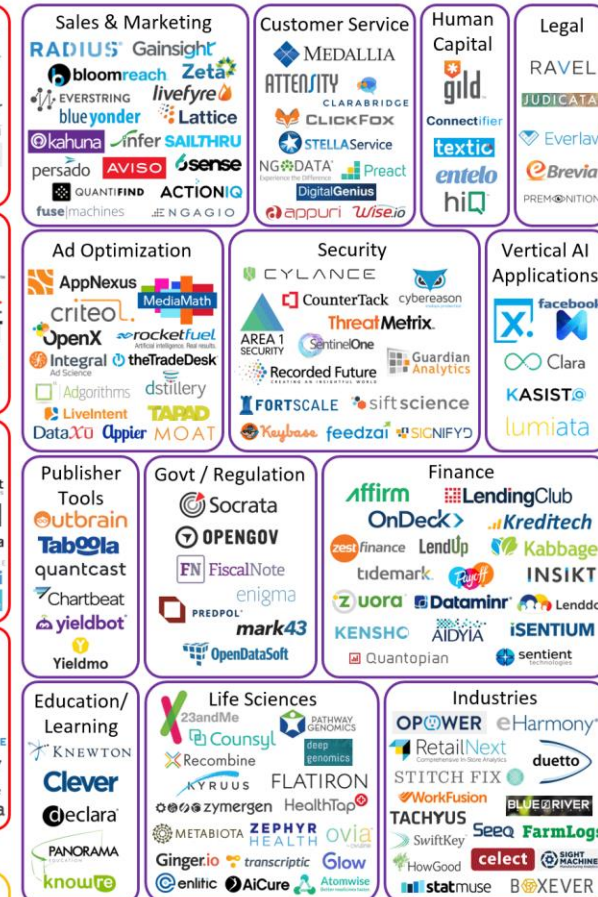
Infrastructure



Analytics



Applications



Cross-Infrastructure/Analytics



Open Source



Data Sources & APIs



Why a Research Data Framework?

- And how do we leverage research data to address global challenges?



RDaF Benefits

- **Increase research integrity** with quality data and improved transparency of the research process
- **Reduce costs and maximize efficiency** by establishing best practices for data management
- **Guide risk management and reduction** through assessment of risk positions and roadmaps for improvement
- **Increase scientific discovery and innovation** with the FAIR principles (Findable, Accessible, Interoperable, Reusable) for better utilization of data

National and International Need

- Data is proliferating at an exponential rate
- Data management is complex and confusing
- Mismanaged data has dire social and economic consequences, including loss of global leadership in critical technical fields
- The U.S. needs a coordinated effort to establish a research data infrastructure, but research data are global in nature so international collaboration / coordination is necessary
- NIST is well-positioned to lead the project; our business is consensus building through being a neutral convener of diverse communities

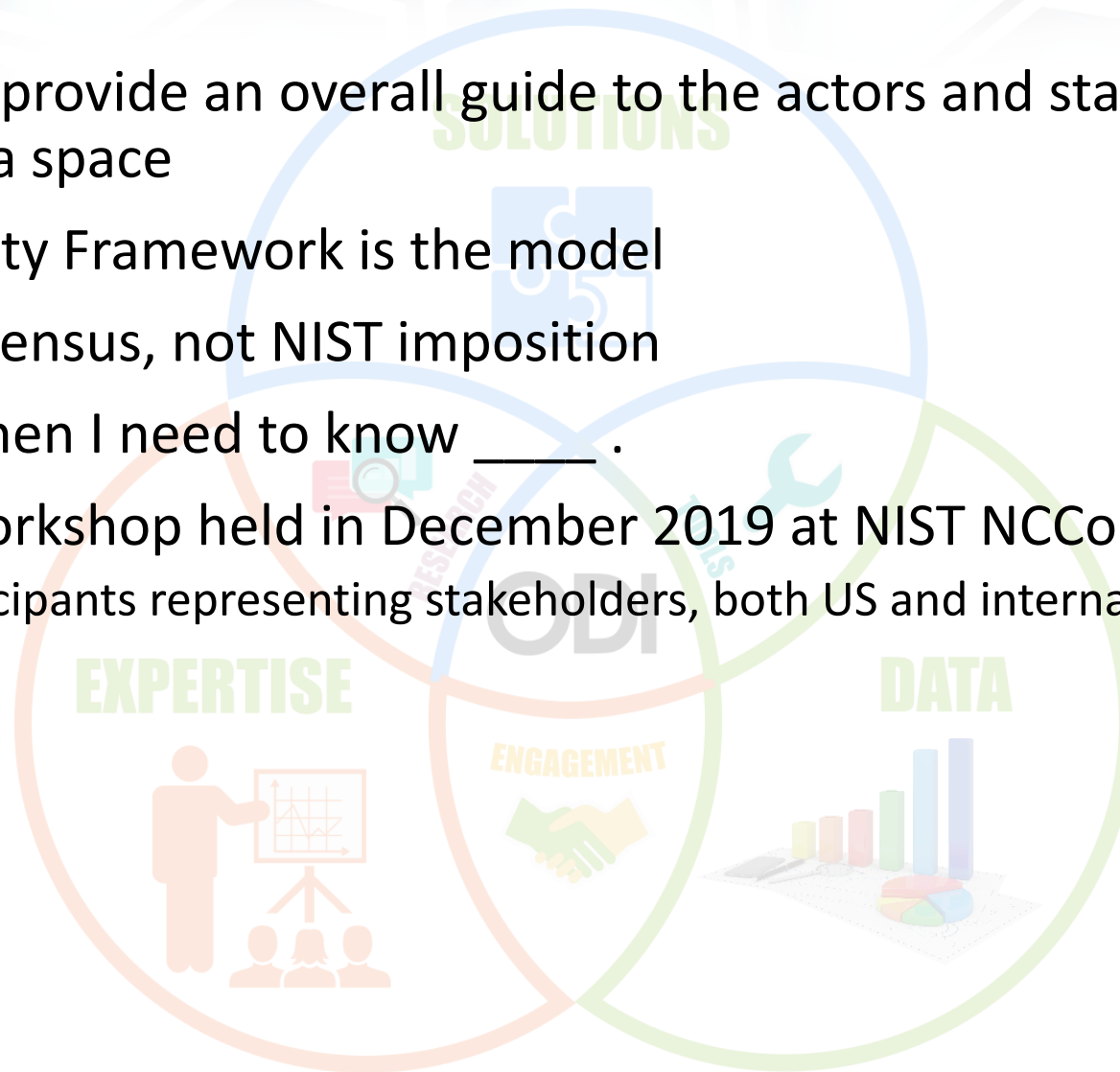
Stakeholders

- Government agencies
- Universities and research libraries
- Data repositories
- Scholarly publishers
- Professional societies
- National and international collaboration organizations (e.g., CENDI, BRDI, CODATA, RDA, WDS, GO-FAIR)
- Standards bodies
- Funders (public and private)
- Industry and the private sector
- Researchers
- General public



Process

- Pilot program to provide an overall guide to the actors and stakeholders in the research data space
- NIST Cybersecurity Framework is the model
- Community consensus, not NIST imposition
- If I am a _____ , then I need to know _____ .
- Initial scoping workshop held in December 2019 at NIST NCCoE
 - 50 invited participants representing stakeholders, both US and international



NIST RDaF Workshop



Research Data Framework

Robert Hanisch
Director, Office of Data
and Informatics
Material Measurement
Laboratory
National Institute of
Standards and
Technology



Bonnie Carroll
Founder & CDO,
Information
International
Associates
Secretary General,
CODATA

RDaF Steering Group



Laura Biven, DOE



Mercé Crosas, Harvard



Josh Greenberg, Sloan



Hilary Hanahoe, RDA



Heather Joseph, SPARC



Barend Mons, CODATA
and GO-FAIR



Beth Plale, NSF



Anita de Waard, Elsevier

Opening Keynotes



Walter Copan, NIST Director



Jim St. Pierre, Deputy Director
Information Technology Laboratory

Workshop Summary

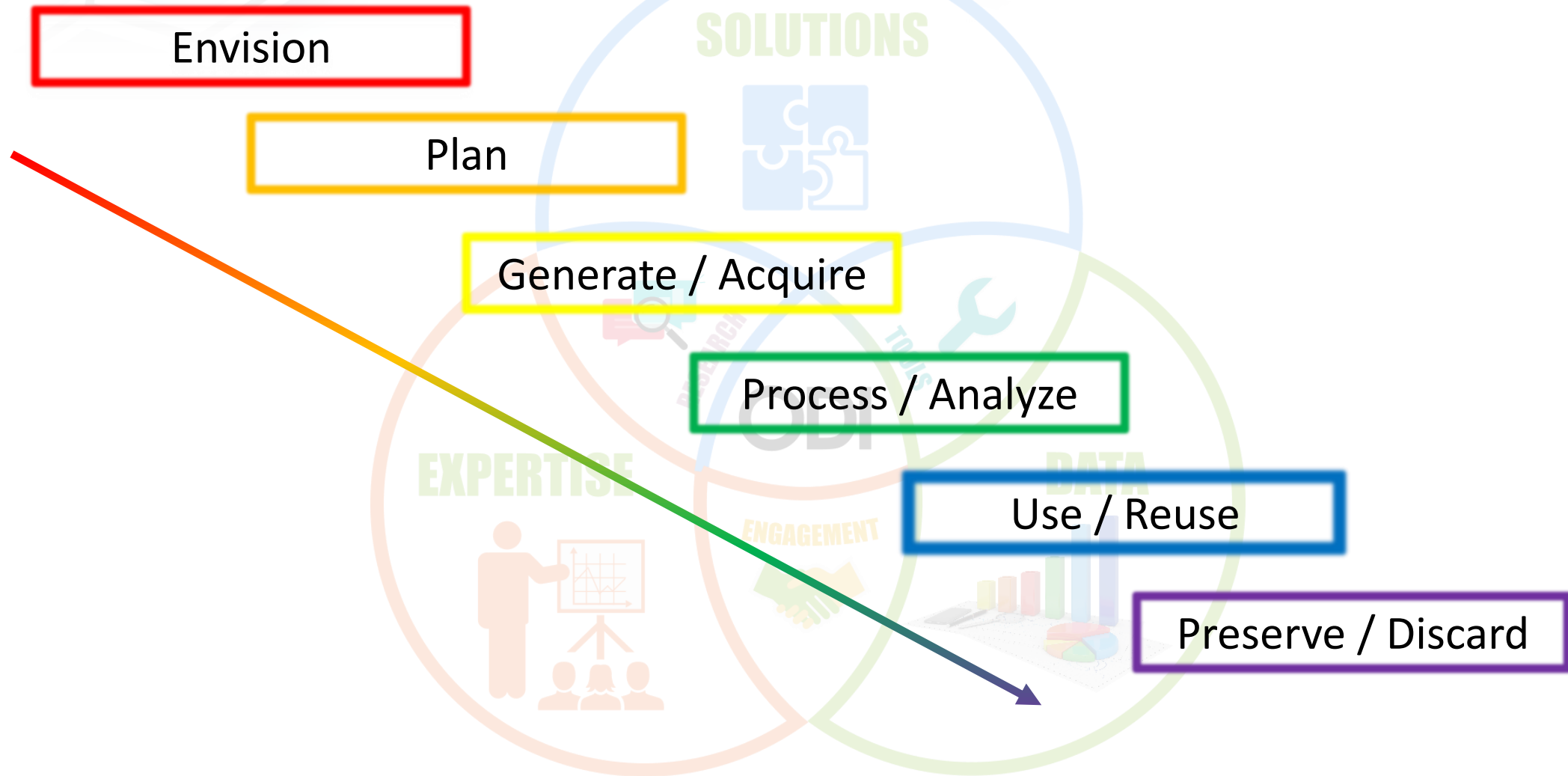
- **Why an RDaF?** An increasing complex data ecosystem with voracious artificial intelligence drivers for trustworthy data requires systematic, intentional data asset management.
- **Purpose:** To optimize use and value of strategic data assets with a coherent data management strategy.
- **Scope:** Covers management of research data created and/or used by any organization. Provides a customizable framework to select and evaluate maturity on key data management factors for each organization.

Workshop Summary (cont'd)

- **Status:** Confirmed support by government agencies, academic organizations, private sector companies, not-for-profit organizations, and international stakeholders.
- **Next Steps:** Management commitment to complete the scoping, pilot testing, and community building for the Framework.
 - Proposed pilots
 - Materials science
 - Universities and research libraries (AAU, APLU, ARL)

Will need cooperation across government to move fully forward with the Framework.

RDaF Structure Based on “Functions”



RDaF Structure

Function

1) Envision

Category

Data governance

*Community
engagement*

Data culture

Reward structure

Subcategory

Data vision, data policy
Data management organization
Data quality, privacy, ethics

Communication, interactions
Cross-domain

FAIR principles
Value of data
Roles and responsibilities

Value of data professionals
Incentives for sharing and re-use

RDaF Structure

Function

2) Plan

Category

Costs

Funding

Data objects

Data management
planning

Subcategory

Cost-benefit analysis
Costs by data lifecycle stage

Direct, overhead, mixed, other

Data (experimental, simulation)
Software, instruments
Publications, presentations

DMPs (intent, update)
Formats, standards

RDaF Structure

<u>Function</u>	<u>Category</u>	<u>Subcategory</u>
3) Generate / Acquire	Sources	In-house, experiment or simulation Collected from external sources
	Experiment	Instruments and their metadata Measurement protocols Data capture and recording
	Simulation	Commercial or custom software Metadata capture and recording
	External sources	Identification, provenance Metadata harvesting
	Data formats	Standards development and/or adoption

RDaF Structure

Function

4) Process / Analyze

Category

Provenance

Data architecture

Software

Publishing, curation

Subcategory

Origin, version, time-stamp
Data copied or derived from other data

Design, security, configuration management
Hosting and storage
On-premise or Cloud

Commercial or custom software
Versions
Stability, resilience, adaptability, maintenance
Workflows, ELNs, LIMs

Processes, tools, stewardship
Metadata

RDaF Structure

Function

Category

Subcategory

5) Use / Reuse

Legal and licenses

Ownership, IP, rights and restrictions
Agreements, permissions
Citation expectations

Data access

Internal, external
APIs
Downloads vs. visiting

Analysis tools

AI/ML
Performance

Impact

Usage tracking, citation

RDaF Structure

Function

6) Preserve /
Discard

Category

Sustainability

Subcategory

Longevity requirements
Who pays?
Orphan data sets

Preservation

Media and media migration
Back-up
Repositories (domain, institutional, general)
Migration between organizations

Retention and disposition

Decision processes
End-of-life (dark archives, deaccession, gravestones)

Status

- Briefed OSTP Subcommittee on Open Science and OSTP Director Kelvin Droegemeier (03/26/2020)
- Discussed pilot concept with AAU, APLU, ARL
- Developed roadmap and structure, vetted with Steering Group
- Preparing to brief NIST upper management
- Will seek ~\$450k to fund two pilots
 - NIST plus other agencies, either \$\$ or in-kind support
 - Professional societies
 - Scholarly publishing community



NIST Frameworks

Framework for Improving Critical Infrastructure Cybersecurity:

<https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf>

NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, September 6, 2019 (Preliminary Draft)

https://www.nist.gov/system/files/documents/2019/09/09/nist_privacy_framework_preliminary_draft.pdf

NIST Big Data Interoperability Framework: Volume 1, Definitions October 2019 Version 3

<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf>

RDaF Summary

- Successful in building community interest and engagement
 - Diverse stakeholders
 - National and international
- Challenges
 - Resources, including our own time within ODI/MML
 - Timeliness: the research data ecosystem is changing rapidly. How to keep pace and assure ongoing updates?
 - Controlling scope and scale
- Strategy for moving forward
 - Confirm NIST support with upper management
 - Start with pilot projects in order to validate approach and re-tune as necessary
 - Collaborate with other federal agencies, professional societies, scholarly publishing community, etc., to garner the necessary resources and take advantage of work in progress

Contact

SOLUTIONS

Robert Hanisch

robert.hanisch@nist.gov

<https://nist.gov/people/robert-hanisch>

